I CLAIM:

1. A method of compressing file-reference data related to information symbols in a source file, comprising steps of:

retrieving location data related to each location of respective ones of the information symbols in the source file;

compressing the location data using a run encoding compression method to construct an encoded data structure; and

storing the encoded data structure in a computer-readable storage medium.

2. The method as claimed in claim 1 wherein the run encoding compression method comprises a step of encoding the location data in the encoded data structure as one or more runs, each run including a first code for representing a first location of the information symbol in a section of the source file encoded by the run, and, if the information symbol occurs more than once in the section of the source file, a second code representing one or more additional locations of the information symbol in the section of the source file.

3. The method as claimed in claim 2 wherein the first code is a binary number representative of a line number in the source file.

4. The method as claimed in claim 3 wherein the second code is a binary string that serves as a bit map

representing one or more additional lines offset from the line number in the source file in which the information symbol occurs at least once.

5. The method as claimed in claim 4 wherein each run has a variable length, and each run further comprises a third code for indicating a length of the second code.

6. The method as claimed in claim 3 wherein each run further includes a fourth code indicating a length of the first code.

7. The method as claimed in claim 1 wherein the source file is a source code file, and the information symbols are source code identifiers.

8. The method as claimed in claim 7 wherein the location data comprises a representation of one or more line numbers in the source code file on which a respective information symbol is referenced.

9. The method as claimed in claim 8 wherein the location data further comprises at least one representation of a column number for each line number represented in the location data and the at least one representation is stored in the computer-readable medium in association with the encoded data structure.

10. The method as claimed in claim 1 further comprising a step of parsing the source file to derive the file-reference data, and the step of retrieving comprises

retrieving the location data from cross-reference line tables built during the step of parsing.

11. The method as claimed in claim 1 further comprising steps of:

compressing the information symbol into a code having a predetermined length; and

storing the code of predetermined length in the computer-readable medium in association with the respective encoded data structure.

12. The method as claimed in claim 1 wherein the data structure is a one of a B-Tree, M-Tree, quad-tree and hashing-based structure.

13. A computer-readable medium containing a file-reference data structure, comprising one or more distinct information symbols and compressed file-reference data representing one or more locations of respective ones of the information symbols in a source file, the compressed file-reference data comprising run encoded location data generated by a run encoding compression method.

14. The computer-readable medium as claimed in claim 13 wherein the encoded data structure field comprises one or more runs, each run comprising a first code for representing a first location of an information symbol in a section of the source file encoded by the run, and, if the reference occurs more than once in the source file, a second code comprising a bitmap

representing one or more additional locations of the reference in the section of the source file.

15. The computer-readable medium as claimed in claim 14 wherein the run further comprises a third code for indicating a length of the second code.

16. The computer-readable medium as claimed in claim 14 wherein the run further comprises a fourth code indicating a length of the first code.

17. The computer-readable medium as claimed in claim 14 wherein the source file is a source code file and wherein the information symbols are source code identifiers.

18. The computer-readable medium as claimed in claim 16 wherein the first code comprises a representation of a line number in the source code file on which a respective source code identifier is referenced.

19. The computer-readable medium as claimed in claim 13 wherein the file-reference data further comprises reference class data representing a use of the information symbol at a location of the reference, the file-reference data structure further comprising, for each of the information symbols and for each of the one or more locations to a reference, a reference class code field stored on the computer-readable medium in association with the respective encoded data structure, said reference class code field for storing reference class data encoded in a predetermined length reference class code.

20. The computer-readable medium as claimed in claim 20 wherein the location data represents one or more locations to a reference to the information symbol in one or more files of information symbols and wherein said locations are associated by file and wherein said file-reference data structure further comprises said first field and said encoded data structure field for each distinct file of information symbols.

21. The computer-readable medium as claimed in claim 13 wherein the index data structure is one of a B-Tree, M-Tree, quad-tree and hashing based structure.

22. An apparatus for compressing file-reference data related to information symbols in a source file, comprising:

means for retrieving location data related to each location of respective ones of the information symbols in the source file;

means for compressing the location data using a run encoding compression algorithm to construct an encoded data structure; and

means for storing the encoded data structure in a computer-readable storage medium.

23. An apparatus as claimed in claim 22 further comprising a fuzzy parser for generating the location data related to each location of respective ones of the information symbols in the source file.

24. An apparatus as claimed in claim 22 wherein the run encoding compression algorithm is adapted to analyze

the location data and construct at least one run associated with each information symbol, each run comprising at least a first code indicating a line number in the source file in which the information symbol appears.

25. An apparatus as claimed in claim 24 wherein the run encoding compression algorithm is further adapted to examine the reference data and construct a binary string that serves as a bitmap offset from the first code to indicate a line location of additional occurrences of the information symbol in the source file.

26. An apparatus as claimed in claim 25 wherein the run encoding compression algorithm is further adapted to determine a length of the run by computing a distance expressed in a total number of lines between a last occurrence of the information symbol in the run and a next occurrence of the information symbol in the source file, and to include the next occurrence in the run if the distance is less than a predetermined threshold and an addition to the bitmap resulting from the inclusion does not make the run longer than a predetermined limit.

27. An apparatus as claimed in claim 26 wherein the apparatus is further adapted to set the predetermined threshold to a number of lines that would cause a length in bytes added to the bit map to exceed an overhead in bytes generated by creating a new run.

28. An apparatus as claimed in claim 26 wherein the apparatus is further adapted to compute the predetermined limit is by computing a capacity of a third code that indicates a length of the second code.